Mapping Internet Connectivity

Craig E. Wills
Worcester Polytechnic Institute

November, 2017

WPI

Introduction

Internet connectivity has become increasingly important as people spend more time living in a virtual world interacting with others who are located literally anywhere on the planet.

However this virtual connectivity is still ultimately layered on physical connectivity where people meet and interact in person with each other.

Interest in physical connectivity led to work on GeoConnected project.

Geographical Connectivity

Use data analysis, mapping and visualization techniques to better understand geographical connectivity, the amount of time to physically travel from one point to another.

Initial focus within the United States, but could be extended worldwide.

Connectivity between two points is not about distance between them, but rather travel on real transportation routes (via automobile and commercial air service) between them.

GeoConnected Project

Gather and analyze data on geographical connectedness at granularity of a *county* as:

1. counties divide the United States geography into relatively small discrete units (albeit of varying size) with data available on population and the county seat (or some other principal city/town) for each county.
2. d3 software is available to visualize the results of analysis.

Project website at `http://geoconnected.cs.wpi.edu/`. Paper submitted for publication.

GeoConnected Methodology

Use data from 3143 FIPS codes in U.S.—basically one per county. Identify a principal city/town within each county (typically county seat).

All travel times between principal cities of two counties are computed based on the smaller of:

1. the time to drive a car from the first principal city to the second principal city, and
2. the time to drive to a nearby airport of the first principal city, plus the time to fly to destination airport nearby the second principal city, plus the time to drive to the second principal city from that destination airport.

Look at GeoConnected site.

Internet Connectivity

Similar two-level infrastructure on the Internet as long-haul, fiber-optic infrastructure (documented in SIGCOMM'15 paper) is analagous to commercial air service network.

Last-mile, ISP connections from residences/businesses to this long-haul network are analagous to regional road structure.

A comparable approach for Internet connectivity would be to determine the time from each point (residence/business) in the United States to every other point.

Appropriate/important for a true point-to-point application—multi-player game?

Beyond Internet Connectivity as Point-to-Point

1. Speedtest service: RTT/upload/download performance to known servers measuring ISP performance.

   Faculty ad for University of Tennessee-Chattanooga, "... Chattanooga also has the fastest internet in the country. Home to the first Gig Internet in the United States, Chattanooga has a 100% fiber network that links every home and business in a 600-square mile area."

2. Perhaps a more appropriate definition of Internet Connectivity is connectivity to Internet infrastructure such a major Internet sites such as Web and Cloud services.

   Does Internet connectivity refer to round-trip time or throughput?

Measurement Approaches

Looking to gather measurements *from* a set of clients *to* a set of servers.

Multiple measurement approaches exist, which define the from-set and to-set.

Approach may also dictate type of measurement (e.g. RTT, throughput) that is available.

PlanetLab (Measurement Infrastructure)

Use existing measurement platform. PlanetLab has been in existence for awhile. Not clear on current status nor geographical diversity of measurement points.

From-Set is a relatively small, fixed set of measurement points.

To-Set is any set of Internet servers.

Both RTT and throughput measurements are possible.

JavaScript/Java App/Mobile App

User volunteer visits a Web site in which JavaScript or a Java App causes browser to retrieve content from any Web server. Similarly a volunteer could download a mobile app. Approach used in How's My Network project at WPI.

From-Set is the set of locations in which volunteers are located at the time of testing.

To-Set is set of Web servers.

RTT and throughput measurements are possible.

Mechanical Turk

Platform pays users to visit a Web site in which JavaScript or a Java App causes browser to retrieve content from a Web site.

From-Set is the set of locations in which volunteers are located at the time of testing.

To-Set is set of Web/Cloud servers.

RTT and throughput measurements are possible.

DNS Clients and Servers

Make use of distributed set of Domain Name System (DNS) recursive and authoritative servers. Recursive servers automatically forward queries to authoritative DNS servers for uncached entries.

From-Set is the set of recursive (public and other) DNS servers willing to respond to DNS queries from arbitrary client.

To-Set is set of authoritative DNS servers.

RTT measurements are possible by computing time difference between client-to-RecursiveServer RTT and client-to-RecursiveServer-to-AuthoritativeServer RTT.

Traceroute

Use *traceroute* to determine routes and "last-mile" RTT for Internet end-points from long-haul network.

From-Set is any Internet end-point in which last-mile time can be determined.

To-Set is same set of points.

If assume that RTTs are symmetric then can estimate connectivity between two Internet points as (last-mile-time + long-haul-time + last-mile-time)

DNS Summary

Seeking diversity of client locations and ideally controllable set of client locations. PlanetLab set is fixed and not growing.

Paid or unpaid set is at mercy of who volunteers—need to provide enough incentive to ensure broad participation. Speedtest has broad participation, but measures last-mile connectivity and not connectivity to Internet services.

Appealing aspect is that the set of recursive/authoritative DNS servers is geographically distributed (residences/businesses/coffee shops/educational institutions). Need to understand if:

1. recursive/authoritative servers for these entities are similiarly distributed—how many are outsourced to third parties?
2. percentage of these recursive servers are willing to respond to arbitrary client requests.

Traceroute Summary

Amount of available information and piecing it together. Use multiple measurement points? Sample traceroute output.

```
traceroute to www.uidaho.edu (129.101.119.250), 30 hops max, 40 byte packets
 1  rtr-flcr1-fuller.inf.wpi.edu (130.215.24.3)  0.347 ms  0.249 ms  0.242 ms
 2  rtr-gpop2-flcr1.inf.wpi.edu (130.215.0.78)  15.959 ms  14.603 ms  14.520 ms
 3  fw-gpop-2-rtr-gpop2.inf.wpi.edu (130.215.0.70)  0.893 ms  0.892 ms  0.834 ms
 4  rtr-gpop2-fw-gpop-2-wan.inf.wpi.edu (207.174.161.5)  1.171 ms  1.110 ms  1.095 ms
 5  18.254.32.41 (18.254.32.41)  5.865 ms  5.803 ms  5.768 ms
 6  i2-re-nox1sumgw1.nox.org (192.5.89.18)  32.841 ms  32.807 ms  32.732 ms
 7  ae-1.4079.sdn-sw.eqch.net.internet2.edu (162.252.70.131)  27.491 ms  27.329 ms  27.390 ms
 8  et-10-0-0.4079.rtsw.chic.net.internet2.edu (162.252.70.132)  28.910 ms  28.850 ms  28.793 ms
 9  ae-3.4079.rtsw.kans.net.internet2.edu (162.252.70.141)  38.563 ms  38.561 ms  38.506 ms
10  et-8-0-0.4079.sdn-sw.denv.net.internet2.edu (162.252.70.10)  48.985 ms  49.037 ms  49.012 ms
11  et-4-1-0.4079.rtsw.salt.net.internet2.edu (162.252.70.9)  58.291 ms  58.274 ms  58.229 ms
12  64.57.28.207 (64.57.28.207)  58.403 ms  58.225 ms  58.315 ms
13   (74.118.19.48)  65.249 ms  65.216 ms  65.231 ms
14   (74.118.19.42)  71.880 ms  71.869 ms  71.959 ms
15  74.118.17.74 (74.118.17.74)  77.465 ms  72.459 ms  72.660 ms
16  lib6500-te5-16.csrv.uidaho.edu (129.101.253.125)  72.321 ms  72.265 ms  72.223 ms
17  lb1-WEB-Sitecore.its.uidaho.edu (129.101.119.250)  72.073 ms  72.129 ms  72.120 ms
```